ENHANCED SUBSCRIBER AUTHENTICATION PROTOCOL

This invention relates to a cryptographic system, and more particularly to an authentication protocol utilizing elliptic curve encryption schemes.

BACKGROUND OF THE INVENTION

In constrained environments where computation power, storage space, and bandwidth are severely limited, as is the case for mobile phones, public-key cryptography was considered inefficient and therefore not a viable option. However, with the advent of faster processors and more efficient public-key cryptographic techniques, such as use of elliptic curve mathematics, traditional obstacles that prohibited use of public-key cryptography have essentially been overcome. Indeed, public-key cryptography has already been incorporated into mobile phones for applications as well as over the air service provisioning. These events allow the wireless industry to exploit the advantages of public-key cryptography in third generation systems.

Mobile phones that communicate over RF networks represent a classic example of the problems facing data security. RF networks are easy to eavesdrop so data sent by a phone can easily be monitored by an adversary and the medium itself prevents data being placed in physical opaque envelopes to ensure secrecy. In fact data sent over RF networks by mobile phones may be subject not just to eavesdropping. It is also possible for an adversary to intercept messages and replace them, delete them, or subvert them. Mobile phones users therefore need diverse security services such as the assurance that data it receives is genuine as well as the assurance that data it sends will remain secret. Data security services needed between a mobile phone and a service provider communicating over an RF network include:

Data confidentiality: Both the mobile phone user and the service provider may want messages they exchange to remain secret. For example, the mobile phone user may want adversaries to be unable to eavesdrop on sensitive calls.

Data integrity: Both the mobile phone user and the service provider may want messages they exchange to remain unaltered. For example, the service provider may want the assurance that the call request it receives specifies the same number that the user dialed so that the call can be completed as dialed.

Data authentication: Both the mobile phone user and the service provider may want to know the origin of data they receive. For example, the service provider may want to know the origin of a call request so that it can decide whether to complete the call.

Non-repudiation: The mobile phone user may wish to send data that is non-repudiable, meaning that the user cannot later deny sending the data. For example, the user may wish to complete a financial transaction such a buying stock over the phone.

Device or entity authentication: Both the mobile phone user and the service provider may want to know who they are communicating with. For example, the service provider may want to check that it is communicating with a paid-up mobile phone user before allowing the user to place calls. Device authentication should prohibit in particular an adversary from replaying the authorization sent by a valid user in order to gain access to the network.

Cryptography is capable of providing all these services. Encryption schemes can be used to provide data confidentiality, message authentication codes (MACs) or signature schemes can be used to provide data integrity and data authentication, and signature schemes can be used to provide non-repudiation. Entity authentication can be provided using more complicated protocols built out of encryption schemes, message authentication codes, and signature schemes.

Currently, security in cellular networks is limited primarily to device authentication. Before allowing a mobile phone network access, the network or service provider authenticates the phone using a protocol based on a message authentication code. The need for additional security services like those listed above has motivated the cellular industry to provide more comprehensive security in future third generation systems.

Authentication in current RF systems consists of device authentication based on symmetric cryptography. The mobile station is provisioned with an Authentication Key, referred to as the A-key, prior to any communication with the cellular network. The A-key is also provisioned in the Authentication Center (AC) of the service provider or home network. The process of provisioning the A-key in the mobile station is part of "service provisioning", during which other mobile station specific information is also provisioned. While there are several ways to provision the A-key, Over-The-Air Service Provisioning (OTASP) is recommended. OTASP uses the Diffie-Hellman protocol to create the A-key concurrently in the AC and the mobile station.

The A-key is then used to create session keys known as SSDs (Shared Secret Data), which are stored in the mobile station and the home network and are used to authenticate

the mobile station. The SSDs are derived by hashing the A-key and other information, such as the mobile station's identity. When a user is roaming in another part of the network, the home network, at the discretion of the service provider, may decide to share the SSD with the serving network to enable the serving network to authenticate the mobile device itself using SSD. Alternatively, the service provider may require the serving network to authenticate the mobile station by checking with the home network each time. Sharing SSDs with the serving network saves signaling traffic between the two networks when the user is roaming but it also requires a degree of trust in the serving network, since knowledge of SSD enables the serving network to impersonate the mobile station.

When a mobile station powers-on, it "registers" with the network. During registration, the mobile station sends its identity to the serving network (assume that the mobile station is roaming) along with an authentication string or MAC (for simplicity, the term MAC is used throughout the rest of the document for authentication string) that is created by hashing SSD, identity information, a random challenge (32-bit number broadcast by the base station), and other information. The serving network queries the user's home network to register the mobile station. The home network, at this point, determines if the SSD is to be shared with the serving network. If so, the SSD is passed to the serving network. The serving network computes the MAC by using the same inputs as the mobile station. If the computed MAC matches the one sent by the mobile station, the mobile station is considered authenticated.

The serving network keeps the SSD associated with the mobile station for the duration of the time that the user is registered in that network. During that time, if the user originates a call (referred to as call origination), the mobile station is again authenticated in the same way as it was for registration, except that dialed digits may be used as additional input to the hash. Once again, the serving network computes the MAC and verifies if the two MACs match, thereby authenticating the mobile station. In addition, if a call is received for the user (referred to as call termination), the same procedure is repeated. In short, authentication is based on calculation of a MAC, which is a hash of SSD, a random challenge from the serving network or base station, and other input, such as mobile station's identity and/or dialed digits.

The Authentication Center in the home network may decide to update the SSD in the mobile station, referred to as SSD update. This is accomplished by sending a request to the mobile station to generate a new SSD. In this scenario, mutual authentication of the mobile

station to the home network and of the home network to the mobile station is performed prior to storage of the newly generated SSD in the mobile station. The authentication of the home network consists of the following: the mobile station sends a random challenge to the AC; the AC computes a MAC using a component of SSD, the random challenge, and other information, and sends it to the mobile station; the mobile station verifies the received MAC with its own computed value.

There are a number of weaknesses with the current authentication system.

It requires the backbone network connecting the home network and the serving network to be very secure. Messages exchanged on this network must be exchanged confidentially; otherwise, an eavesdropper monitoring this channel can impersonate any active mobile stations.

It imposes high security requirements on the Authentication Center of each service provider. Maintaining the confidentiality of the A-key database at the Authentication Center is essential; otherwise, anyone who learns the contents of the database can impersonate any mobile station at any time. This problem is escalated by the fact that there is no effective disaster recovery mechanism in the event of Authentication Center compromise.

There are security concerns over SSD sharing. If the home network decides to share SSDs with serving networks, this enables the serving network to impersonate mobile stations.

The CAVE algorithm, which is used to provide authentication, itself has security concerns. CAVE has not been published and has not received widespread scrutiny by the cryptographic community. Compromise of CAVE could cause embarrassment for the cellular community, which has already been hurt by the use of unpublished algorithms, such as CMEA.

There are efficiency concerns. A large amount of communication is required on the backbone network linking the home network and the serving network. This communication is substantially increased if the home network is not sharing SSDs with the serving network since now the serving network must communicate with the home network each time it wants to authenticate the mobile station.

Most importantly, the current system does not provide sufficient security services. Device authentication provides limited security to the network since there remains the possibility that an adversary can hijack service after device authentication has been performed. Furthermore, the limited deployment of data privacy services presents a major

problem since it means users are wary of placing sensitive calls over the cellular network. This issue will become particularly important in the future if the cellular industry wants to support advanced features like internet browsing and over the air financial transactions.

Global roaming, one of the most promising features of third generation systems, will heighten many of these concerns.

The deployment of a third generation system affords the cellular industry an opportunity to address the deficiencies of the current authentication system. The third generation authentication system therefore needs to meet the following requirements:

Minimization of computation time required by mobile stations for generation of appropriate keys on each access. Since authentication is used for every call, performance is an important consideration. Security should not affect the service being offered to the end-user negatively.

Ability to provide non-repudiation. This is extremely useful in services that are expected to drive deployment of third generation systems.

Minimization of extra network infrastructure. Since third generation is a migration from second generation systems, it is important to take advantage of the current infrastructure in place, where possible.

Scalability. As more and more cellular systems are brought into service each year and with many carriers aiming to provide worldwide roaming, third generation ESA and ESP should provide for the ability to scale without imposing additional costs on carriers.

SUMMARY OF THE INVENTION

This invention seeks to provide a system for subscriber authentication in a network that obviates and mitigates the disadvantage of current systems by utilizing a combination of public-key cryptography and symmetric cryptography.

It is an object of the present invention to provide for mutual authentication of network and mobile station.

In accordance with this invention there is provided a method of authenticating a pair of correspondents in a communication system, said method comprising the steps of:

Exchanging cryptographic keys between said corespondents, said exchange being based on a public key mutual authentication scheme; and

using said keys for encrypting data in a symmetric-key data exchange.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features of the preferred embodiments of the invention will become more apparent in the following detailed description in which reference is made to the appended drawings wherein:

Figure 1 is a schematic diagram of an RF communication system;

Figure 2 is a schematic diagram showing an authenticated key establishment protocol according to an embodiment of this invention; and

Figure 3 is a schematic diagram showing a secure data exchange according to an embodiment of this invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring to figure 1, a schematic diagram of a communication system is shown generally by numeral 10. The system 10 comprises a mobile station 12, a base station 14, a service provider 16, all of which communicate via a network 18. The network may be RF, cellular, satellite communication or the like.

A third generation system according to an embodiment of the patent invention is based around the same events as the current systems: service provisioning, mobile registration, call origination, call termination, and data exchange. However there are two fundamental differences between the subject system and the current system. First, the proposed system performs registration using a protocol that provides mutual authentication and key establishment and is based on public-key cryptography. Second, the subject system also includes provision to secure data exchange.

Before a mobile station can begin using the cellular network, service provisioning must take place. Elliptic curve parameters are embedded in the mobile station at manufacture time. During service provisioning, the mobile station selects at random an elliptic curve private key and computes the corresponding elliptic curve public key. It now exchanges this public key and its identity confidentially with its service provider who stores the information in its Authentication Center. A variety of mechanisms that can be used by the mobile station and the service provider to exchange the mobile station's public key.

Service provisioning refers to the process of establishing a secure key pair in the mobile station and exchanging all the necessary data between the mobile station and the service provider so that the mobile station is ready to communicate with the network. From

a cryptographic perspective, the salient points of this process are key generation and public key exchange.

Key generation can be performed efficiently by the mobile station at any time prior to service provisioning. One approach is to embed a random seed in the mobile station during manufacture and use this seed to generate the private key. This means that the security of the mobile's private key is not based solely on randomness generated by the mobile station. This is desirable because it is notoriously hard to generate randomness on constrained devices.

Having generated a secure key pair at the mobile station, there are many ways to perform public key exchange. It is likely that any deployment of the system would employ a variety of methods. Some of the possibilities are listed below.

Manual exchange during activation at a distributor outlet. The public key of the mobile station could be transmitted securely from the outlet to the Authentication Center of the service provider using a dial-up connection.

Exchange at manufacture time. The manufacturer of the mobile station could retrieve the public key during manufacture and then transmit the public key securely to the service provider when the mobile station user requests service.

Over-the-air exchange. The mobile station and the service provider could exchange the public key over the air when the mobile user requests service. In this case the exchange could be secured using a password established between the user and the service provider, or using a public key of the manufacturer embedded in the mobile station at manufacture time.

Any of these mechanisms would likely provide sufficient security if implemented properly. Over-the-air activation is particularly appealing because it is potentially transparent to the mobile station user.

Service provisioning is completed by establishing an account associated with the mobile station so that the mobile station user can be charged for network service.

To obtain service within a serving network, the mobile station powers on and registers with the network. As in current cellular systems, registration consists of a protocol completed between the base station and the mobile station. However, the subject registration mechanism accomplishes considerably more than just device authentication of the mobile station. It provides mutual authentication of both the mobile station to the base station and the base station to the mobile station and in addition establishes session keys, which are then used to secure future communications between the base station and the

mobile station during the session. Authentication and key exchange are supplied by a public-key based protocol outlined below.

The base station enables registration by sending a short-lived elliptic curve public key along with its identifier to the mobile station. This information can either be broadcast on the overhead channel or it can be sent to an individual mobile station in response to a registration request from the mobile station.

In response, the mobile station combines the short-lived base station public key with its own private key and generates two shared secret keys using the elliptic curve Diffie-Hellman method. The first of these two keys is used as a MAC key to authenticate the mobile station to the base station and the base station to the mobile station, and the second is used to establish secret session keys. It then chooses a random challenge, computes an authentication string (i.e., MAC) using the established MAC key, and sends its identity, the random challenge, and the MAC to the base station in order to register.

On receiving the registration request from the mobile station, the base station first contacts the mobile station's service provider, and requests the mobile station's public key. Using the mobile station's public key and its own short-lived private key, the base station also computes two shared secret keys using the elliptic curve Diffie-Hellman method. It uses the first key to check the MAC it received from the mobile station. If this check is successful, the base station registers the mobile station and establishes the mobile station's location, calculates two session keys, and computes a MAC which it uses to authenticate itself to the mobile station. It sends the MAC and the encrypted keys to the mobile station.

Finally, the mobile station checks the validity of the MAC it received from the base station and in turn calculates two session keys. This completes the registration protocol having authenticated both the mobile station to the base station and the base station to the mobile station and established session keys that can be used to secure future communications.

In summary, registration consists of a protocol performed by the mobile station and the base station, which provides mutual authentication and session key establishment. Mobile station authentication is based on its knowledge of its private key. Base station authentication is based on its knowledge of the mobile station public key.

After registration, when the mobile station and the base station wish to exchange data, they use the session keys established during registration to secure the exchange. The same cryptographic mechanism can be used to secure data whether it is to be exchanged on

the control channel or the voice channel, and even if the data is a call origination request from the mobile station to the base station or a call termination request (i.e. a page) from the base station to the mobile station.

The sender takes the data and, if privacy is on, encrypts the data using the first session key. It appends to the result a counter indicating the number of messages that have been exchanged so far, and a direction flag indicating whether the data is being sent from the mobile station to the base station or from the base station to the mobile station. It then MACs the resulting string using the second session key and sends the encrypted data along with the MAC.

This process is described in detail in below.

It is worth noting that the subject system is in many ways less complicated than heretofore authentication systems. The use of public-key cryptography means there is no need for two-tier key management using A-keys and SSDs. This is because in the subject system service providers do not need to share information with serving networks which enables the serving networks to impersonate the mobile station. The use of session keys to secure data exchange means that it is not necessary to re-authenticate the mobile station during call origination and call termination. Instead call origination and call termination can be secured just like any other data exchange between the mobile station and the base station using the session keys.

Registration is performed each time a phone powers on and attempts to establish service in a serving network. Registration consists of an authenticated key establishment protocol which provides authentication of the mobile station to the base station and of the base station to the mobile station, as well as supplying session keys to the mobile station and the base station which they will use to secure future communications.

The key establishment protocol employs a number of cryptographic primitives: the elliptic curve Diffie-Hellman key agreement with SHA-1 based key derivation as described in ANSI X9.63, and the SHA-1 based HMAC message authentication code as described in N. Koblitz. *Elliptic curve cryptosystems*. Mathematics of Computation, 48, pages 203-209. 1987.

The following notation is used in the following description:

Base Station = BS

Base Station's identifying information = BS_ID

Base Station's short-lived private key = b

Base Station's short-lived public key = bP

Cofactor of elliptic curve = t

Direction indicator (MS-originated or BS-originated) = Dir

Elliptic curve parameters = E

Encryption key for key establishment = k'

Encryption key used to encrypt subsequent communications = c'

Generator point on elliptic curve = P

HMAC on M under key k = $MAC_k\{M\}$

Key derivation function based on SHA-1 = KDF

Message Authentication Code = MAC

Message counter = Count

MAC key for key establishment = k

MAC key used to MAC subsequent communications = c

Mobile station = MS

Mobile Station's identifying information = MS_ID

Mobile Station's private key = m

Mobile Station's public key = mP

Mobile Station's random challenge = MS_RC

Symmetric Encryption of M using k' = $ENC_{k'}(M)$

Symmetric Decryption of C using k' = $SKD_{k'}(C)$

Referring to figure 2, a cryptographic protocol according to an embodiment of the present invention is shown generally by numeral 20.

The base station sends its short-lived public key bP and its identity BS_ID to the mobile station. Like the base station random challenge in the current authentication system, bP should either be unique to a single registration (when it is sent upon request) or should be changed frequently.

The mobile station calculates tmbP from bP using its private key m, checks tmbP≠O, and generates two keys k and k' from tmbP using SHA-1: k,k' = KDF{tmbP}. The use of the cofactor t prevents small subgroup attacks.

The mobile station generates a random challenge MS_RC. It computes a MAC under key k on 2,MS_ID,BS_ID,MS_RC,bP: $MAC_k\{2,MS_ID,BS_ID,MS_RC,bP\}$.

The mobile station sends a registration request message to the base station consisting of MS_ID, MS_RC, and the MAC computed as above: $MAC_k\{2,MS_ID,BS_ID,MS_RC,bP\}$. The base station contacts the mobile station service provider, and retrieves the public key of the mobile station.

The base station calculates tmbP from mP and its short-lived private key b, checks tmbP≠O, and generates the keys k and k' from tmbP using SHA-1: k,k' = KDF{tbmP}.

The base station computes the MAC on the same information as the mobile station and compares its computed value with the value sent by the mobile station. If the value matches, the mobile station is authenticated by showing that that it knows its private key m. The base station updates the location of the mobile station, and computes a MAC under key k on 3,BS_ID,MS_ID,bP,MS_RC: $MAC_k\{3,BS_ID,MS_ID,bP,MS_RC\}$.

The base station sends a registration acknowledge message to the mobile station consisting of the MAC computed above: $MAC_k\{3,BS_ID,MS_ID,bP,MS_RC\}$.

The mobile station computes the MAC on the same information as the base station and compares its computed value with the value sent by the base station. If the value matches, the base station is authenticated by showing that it knows the mobile station public key mP.

Both the mobile station and the base station calculate session keys c and c' from the shared secret key k', the base station short-lived public key bP, and the mobile station challenge MS_RC: c,c' = KDF{k',bP,MS_RC}. The session keys c and c' are stored and used to secure future communications in this session between the mobile station and the base station.

This protocol achieves mutual authentication and session key establishment. The session keys established should be sufficient to secure all future communications in this session between the mobile station and the base station until either the base station or the mobile station issues a session termination request. Session termination may occur either because the mobile station roams away from the base station or because the mobile station powers off. Alternatively the base station may choose to terminate the session at any time and require the mobile station to register again.

A variant of the protocol may also be of use in third generation systems. Instead of achieving base station authentication by keeping the mobile station public key secret within the network, base station authentication could be achieved using a long-lived base station public key stored in a certificate issued to the base station by the service provider. In this

variant, the public key of the service provider would be downloaded into the mobile station during service provisioning. The base station would send its certificate along with a random challenge to the mobile station during the first flow of the protocol. The protocol would then proceed as before. The advantage of this approach is that it does not require the network to maintain the secrecy of mobile station public keys. The disadvantage is that it requires service providers to act as Certification Authorities and issue certificates to base stations, and it requires increased computation by mobile station since they must verify the certificate of the base station during registration.

Once the mobile station and base station have authenticated each other and established the session keys during the registration process, future communications, such as call origination, call termination, etc., are secured using the session keys. The mechanism proposed here provides a combination of data confidentiality, data integrity, and data authentication.

The mechanism is illustrated in Figure 3 and described below. It employs HMAC using SHA-1 for authentication and integrity and a generic cipher for privacy.

Suppose the mobile station and the base station want to exchange some data m and that this is the i^{th} message they have exchanged during this session. Then the sender first encrypts the message under key c' using the cipher. Then the sender appends to the encryption a counter whose value is i indicating that this is the i^{th} message exchanged during the session and a direction flag which is a single bit indicating whether the message is being sent from the mobile station to the base station or from the base station to the mobile station. The sender then MACs the resulting string using HMAC with SHA-1 under key c. The encrypted data and the MAC are sent to the receiver.

When the encrypted data and the MAC are received, the receiver first checks the authenticity of the message. They append the appropriate counter value and direction flag to the encrypted data and recalculate the MAC value using c. If the MACs are the same, they have confirmed the authenticity of the message. The recipient then recovers the data itself by decrypting the encrypted data using the cipher under key c'.

This process ensures the authenticity, confidentiality, and integrity of the data. Use of the direction indicator prevents an attack where an active adversary bounces a message back to the sender. Use of the message counter prevents an active adversary from reordering messages.

This process is executed whenever the mobile station and the base station want to exchange data. The same process is used for call origination requests, call termination requests, voice, and any other data exchange during the session.

There are two variations on this process that may be used:

If privacy is off, then the sender does not encrypt the data m. Instead, the sender includes the message in the clear in the MAC calculation, thus sending the following to the recipient: m, $MAC_c\{Dir,Count,m\}$.

If privacy is on and it is deemed that the cipher being used is capable of providing authenticity as well as privacy, then the sender may use the cipher both for authenticity and privacy. In this case, the sender sends the following to the recipient: $ENC_{c'}\{Dir,Count,m\}$. Some systems often assume that block ciphers like DES are capable of providing both authenticity and confidentiality. This variant can save the computational resources of the sender and the recipient since only one symmetric operation is required.

In the event that different privacy algorithms are being used on the control channel and the traffic channel, three session keys - c, c', and c'' - are established during registration instead of two. Messages are now exchanged securely just as described above, except that c' is used to encrypt m if m is being sent on the control channel, and c'' is used to encrypt m if m is being sent of the traffic channel.

Although the invention has been described with reference to certain specific embodiments, various modifications thereof will be apparent to those skilled in the art without departing from the spirit and scope of the invention as outlined in the claims appended hereto.